MEA PAAP
Standards Validation

Overview of Meeting

On December 5, 2003 a meeting was held to validate the standards established for the MEA PAAP. The meeting was conducted in three concurrently running sessions, one for each content area (ELA, Math, Science). Each session ran for one full day.

The standards validation methodology implemented was the same for all content areas. To help ensure consistency of procedures between panels, each panel was led through the standards validation process by trained facilitators from Measured Progress. The methodology employed is based on a systematic review of student work and collaborative panel discussion. This process was designed specifically to meet the needs of the MEA PAAP. An overview of this method is described below.

Overview of Process

This section of the report provides an overview of the standards validation process as it was implemented for the MEA PAAP. The process is divided into the following three stages, each with several constituent tasks.

- Tasks Completed Prior to the Standards Validation Meeting
    1. Creation of Performance Levels and Performance Level Definitions
    2. Preparation of Materials for Panelists
    3. Preparation of Presentation Materials
    4. Selection of Panelists
- Tasks Completed During the Standards Validation Meeting
    5. Orientation
    6. Review Assessment Materials
    7. Review Performance Level Definitions
    8. Review and Ratings of Portfolios
- Tasks Completed After the Standards Validation Meeting
    9. Analysis and Review Panelists ratings and evaluations
    10. Prepare Report of Meeting

Tasks Completed Prior to the Standards Validation Meeting.

1. Creation of Performance Levels and Performance Level Definitions

The MEA PAAP has previously defined 5 performance levels. These performance levels, how they are attained for each rubric level, and the performance level definitions are as follows:

Levels of Assistance	Performance Level			
	1	2	3	4
1=Support	A	Em	P	T
2=Prompting	A	Em	P	T
3=Limited Prompting	A	Em	T	Ex
4=Independent	A	Em	T	Ex

A = Attempting: *Scores indicate that the student is at the initial stages of development of knowledge and skills defined at the PAAP Rubric level identified.*

Em = Emerging: *Scores indicate that the student has developed basic knowledge and skills related to some of the components of the standards defined in the PAAP Rubric level identified.*

P = Progressing: *Scores indicate that the student has developed partial knowledge and skills related to the standards as defined in the PAAP Rubric level identified, and is progressing towards meeting the standards for the PAAP Rubric Level identified.*

T = Transitioning: *Scores indicate that the student has consistently met the standards for the PAAP Rubric level identified and is moving towards the next PAAP Rubric Level.*

Ex = Exceeding: *Scores indicate that the student has an in-depth understanding of the knowledge and skills for the PAAP Rubric Level identified, and can consistently demonstrate these understandings with very little, if any, assistance. Evidence for some of the Content Area standards may be scored at a higher PAAP Rubric Level than is identified.*

As can be seen by the table above, the classification of portfolios into each of these proficiency levels is highly related to the scoring of the portfolios. This relationship reflects the closeness between the standards assessed and the evidence submitted. In a sense the original standard setting has been built into the structure of the portfolios (i.e., the kind of evidence that can be submitted and the way in which that evidence is scored). The purpose of the standards validation meeting was to validate that the standards applied to the portfolios produces fair and proper proficiency level classifications of the submitted portfolios.

2. Preparation of Materials for Panelists

The following materials were assembled into folders for presentation to the panelists at the standard setting meeting:

A. Meeting Agenda
B. Confidentiality Agreement
C. Performance Level Definitions
D. Scoring Rubrics
E. Selected Portfolios

F. Evaluation Forms

3. Preparation of Presentation Materials

The presentation materials (powerpoint, overhead slides, etc…) used in the opening session were prepared prior to the meeting. A copy of these materials is included as an Appendix in this report.

4. Selection of Panelists

Panelists were selected prior to the standard setting meeting. The goal of panelist selection was to have 15 panelists included on each panel. Table 1 below shows the number of panelists on each panel.

Table 1
N umber of Standards Validation
Panelists for each Content Area

Content	#
Math	12
ELA	14
Science	12

Tasks Completed During the Standards Validation Meeting.

5. Orientation

The Standards Validation meeting began with a general orientation session attended by panelists for all content areas. The purpose of this session was to provide an introduction to the issues of standard setting, draw the distinction between standard setting and standards validation, and to explain the activities that would occur during the standards validation meeting. The opening session also provided an introduction to the portfolios, how they were scored, and the rationale for the standards that were set on the portfolios. Throughout the opening session, panelists were encouraged to ask clarifying questions.

After this large-group session, the panelists were assembled into their smaller content specific groups.

6. Review Assessment Materials

In this step of the process, each panelist reviewed the requirements for assembling and submitting a completed portfolio. This was to allow panelists to become familiar with the components of the portfolio and the various forms submitted evidence can take.

7. Review Performance Level Definitions

In the next step of the process, panelists reviewed the performance level definitions that were developed. It was explained to the panelists that it is their task to assess whether the portfolios assigned to each

proficiency level exhibit the knowledge and skills required of that level. To accomplish this task, panelists will need a solid understanding of the proficiency level definitions.

8. Review and Ratings of Portfolios

After the panelists reviewed the proficiency level definitions, they engaged in an activity in which they compared the actual student portfolios against the proficiency level definitions. It was intended that there be three portfolios provided at each proficiency level for each content group, resulting in a total of 15 portfolios: 3 Attempting, 3 Emerging, 3 Progressing, 3 Transitioning, and 3 Exceeding. Table 2 below provides the actual number of portfolios used in the standards validation process. The discrepancy between the targeted and actual number of portfolios resulted from excluding some portfolios at the time of the standards validation meeting. Portfolios were excluded because of incomplete or inappropriate pieces of evidence included in the portfolio (i.e., mathematics evidence in what was designated as a Science portfolio).

Table 2

Number of Portfolios Evaluated
at each Proficiency Level by Content Area

Proficiency Level	Math	ELA	Science
Attempting	2	3	2
Emerging	3	2	3
Progressing	3	3	3
Transitioning	3	3	3
Exceeding	3	3	3

Panelists were led through two rounds of rating the portfolios. In the first round, panelists were to independently evaluate each portfolio and classify it into the performance level they believed it most closely matched in terms of the knowledge, skills, and abilities required by the proficiency level and those exhibited in the portfolio. In the second, round panelists discussed their initial ratings and were told the original classification of each portfolio. Panelists then were provided the opportunity to revise their initial ratings. The results of the panelists' ratings for each content are provided in Tables 3-5.

Each of Tables 3-5 contains the results for both the panelists' initial and final ratings. The initial ratings are in the upper portion of the table and the final ratings are in the lower portion of the table.

For purposes of this meeting, only those portfolios in Rubric level 1 were examined. The reason for this is that the vast majority of the portfiolios submitted were scored at this rubric level, and the purpose of this meeting was not to establish the standards, but to validate that the established standards represented a fair classification of student work.

Table 3										
Mathematics: Results of Panelists' Ratings of Each Portfolio										
Round	Portfolio Number	Proficiency Level					Summary Statistics			
		Attempting	Emerging	Progressing	Transitioning	Exceeding	n	Mean	SD	Category Avg
Initial Rating	1	75.0	16.7	0.0	8.3	0.0	12	1.42	0.90	1.71
	2	41.7	33.3	8.3	16.7	0.0	12	2.00	1.13	
	3	33.3	41.7	16.7	8.3	0.0	12	2.00	0.95	2.83
	4	0.0	8.3	41.7	50.0	0.0	12	3.42	0.67	
	5	18.2	9.1	27.3	36.4	9.1	11	3.09	1.30	
	6	0.0	16.7	66.7	16.7	0.0	12	3.00	0.60	2.72
	7	0.0	16.7	50.0	25.0	8.3	12	3.25	0.87	
	8	41.7	33.3	16.7	8.3	0.0	12	1.92	1.00	
	9	0.0	9.1	45.5	27.3	18.2	11	3.55	0.93	3.06
	10	0.0	20.0	40.0	40.0	0.0	10	3.20	0.79	
	11	30.0	30.0	20.0	10.0	10.0	10	2.40	1.35	
	12	0.0	0.0	30.0	40.0	30.0	10	4.00	0.82	3.90
	13	0.0	0.0	20.0	50.0	30.0	10	4.10	0.74	
	14	0.0	10.0	20.0	70.0	0.0	10	3.60	0.70	
Final Rating	1	91.7	0.0	8.3	0.0	0.0	12	1.17	0.58	1.58
	2	33.3	41.7	16.7	8.3	0.0	12	2.00	0.95	
	3	50.0	25.0	25.0	0.0	0.0	12	1.75	0.87	2.91
	4	0.0	8.3	33.3	58.3	0.0	12	3.50	0.67	
	5	0.0	18.2	27.3	36.4	18.2	11	3.55	1.04	
	6	0.0	16.7	75.0	8.3	0.0	12	2.92	0.51	2.67
	7	0.0	16.7	50.0	25.0	8.3	12	3.25	0.87	
	8	41.7	41.7	8.3	8.3	0.0	12	1.83	0.94	
	9	0.0	8.3	41.7	33.3	16.7	12	3.58	0.90	3.14
	10	0.0	16.7	33.3	41.7	8.3	12	3.42	0.90	
	11	27.3	36.4	18.2	9.1	9.1	11	2.36	1.29	
	12	0.0	0.0	8.3	66.7	25.0	12	4.17	0.58	4.06
	13	0.0	0.0	9.1	54.5	36.4	11	4.27	0.65	
	14	0.0	0.0	27.3	72.7	0.0	11	3.73	0.47	

Table 4

ELA: Results of Panelists' Ratings of Each Portfolio

Round	Portfolio Number	Performance Level					Summary Statistics			
		Attempting	Emerging	Progressing	Transitioning	Exceeding	n	Mean	SD	Category Avg
Initial Rating	1	78.6	21.4	0.0	0.0	0.0	14	1.21	0.43	1.24
	2	92.9	0.0	0.0	7.1	0.0	14	1.21	0.80	
	3	85.7	0.0	14.3	0.0	0.0	14	1.29	0.73	
	4	0.0	78.6	14.3	7.1	0.0	14	2.29	0.61	2.61
	5	7.1	21.4	42.9	28.6	0.0	14	2.93	0.92	
	6	0.0	0.0	50.0	28.6	21.4	14	3.71	0.83	2.98
	7	7.7	46.2	38.5	7.7	0.0	13	2.46	0.78	
	8	7.7	23.1	61.5	7.7	0.0	13	2.69	0.75	
	9	0.0	23.1	61.5	15.4	0.0	13	2.92	0.64	3.38
	10	0.0	0.0	16.7	50.0	33.3	12	4.17	0.72	
	11	0.0	8.3	75.0	16.7	0.0	12	3.08	0.51	
	12	0.0	18.2	9.1	45.5	27.3	11	3.82	1.08	3.52
	13	0.0	30.0	50.0	20.0	0.0	10	2.90	0.74	
	14	0.0	0.0	40.0	40.0	20.0	10	3.80	0.79	
Final Rating	1	92.9	7.1	0.0	0.0	0.0	14	1.07	0.27	1.12
	2	100.0	0.0	0.0	0.0	0.0	14	1.00	0.00	
	3	85.7	0.0	14.3	0.0	0.0	14	1.29	0.73	
	4	0.0	85.7	14.3	0.0	0.0	14	2.14	0.36	2.64
	5	0.0	7.1	71.4	21.4	0.0	14	3.14	0.53	
	6	0.0	0.0	50.0	28.6	21.4	14	3.71	0.83	2.95
	7	0.0	69.2	30.8	0.0	0.0	13	2.31	0.48	
	8	0.0	23.1	76.9	0.0	0.0	13	2.77	0.44	
	9	0.0	23.1	76.9	0.0	0.0	13	2.77	0.44	3.32
	10	0.0	0.0	16.7	58.3	25.0	12	4.08	0.67	
	11	0.0	0.0	83.3	16.7	0.0	12	3.17	0.39	
	12	0.0	0.0	9.1	63.6	27.3	11	4.18	0.60	3.64
	13	0.0	18.2	72.7	9.1	0.0	11	2.91	0.54	
	14	0.0	0.0	36.4	45.5	18.2	11	3.82	0.75	

Table 5

Science: Results of Panelists' Ratings of Each Portfolio

Round	Portfolio Number	Performance Level					Summary Statistics			
		Attempting	Emerging	Progressing	Transitioning	Exceeding	n	Mean	S.D.	Category Avg
Initial Rating	1	91.7	8.3	0.0	0.0	0.0	12	1.08	0.29	1.25
	2	66.7	25.0	8.3	0.0	0.0	12	1.42	0.67	
	3	0.0	50.0	41.7	8.3	0.0	12	2.58	0.67	2.72
	4	8.3	16.7	66.7	8.3	0.0	12	2.75	0.75	
	5	0.0	25.0	66.7	8.3	0.0	12	2.83	0.58	
	6	8.3	16.7	25.0	33.3	16.7	12	3.33	1.23	2.51
	7	33.3	50.0	16.7	0.0	0.0	12	1.83	0.72	
	8	36.4	18.2	18.2	27.3	0.0	11	2.36	1.29	
	9	0.0	50.0	30.0	20.0	0.0	10	2.70	0.82	3.34
	10	0.0	11.1	55.6	11.1	22.2	9	3.44	1.01	
	11	0.0	0.0	30.0	50.0	20.0	10	3.90	0.74	
	12	11.1	0.0	11.1	22.2	55.6	9	4.11	1.36	3.91
	13	14.3	14.3	0.0	57.1	14.3	7	3.43	1.40	
	14	0.0	0.0	16.7	50.0	33.3	6	4.17	0.75	
Final Rating	1	91.7	8.3	0.0	0.0	0.0	12	1.08	0.29	1.17
	2	75.0	25.0	0.0	0.0	0.0	12	1.25	0.45	
	3	0.0	50.0	41.7	8.3	0.0	12	2.58	0.67	2.81
	4	8.3	25.0	58.3	8.3	0.0	12	2.67	0.78	
	5	0.0	0.0	83.3	16.7	0.0	12	3.17	0.39	
	6	8.3	8.3	25.0	16.7	41.7	12	3.75	1.36	2.60
	7	41.7	58.3	0.0	0.0	0.0	12	1.58	0.51	
	8	9.1	45.5	36.4	9.1	0.0	11	2.45	0.82	
	9	0.0	30.0	40.0	30.0	0.0	10	3.00	0.82	3.66
	10	0.0	0.0	66.7	11.1	22.2	9	3.56	0.88	
	11	0.0	0.0	10.0	40.0	50.0	10	4.40	0.70	
	12	0.0	0.0	0.0	11.1	88.9	9	4.89	0.33	4.36
	13	0.0	14.3	0.0	85.7	0.0	7	3.71	0.76	
	14	0.0	0.0	16.7	33.3	50.0	6	4.33	0.82	

The information in Tables 3-5 provides the percentage of panelists that categorized each portfolio into each performance level for both the initial and final rounds of standards validation. For example, based on the initial ratings of Portfolio #1, 75% percent of the panelists categorized that portfolio as *Attempting*, while 16.7% classified Portfolio #1 as *Emerging*, and 8.3% classified it as *Transitioning*. In the final round of Ratings Portfolio #1 was classified as *Attempting* by 91.7% of the panelists and as *Progressing* by 8.3% of the panelists.

The shaded cells in Tables 3-5 represent the classification of the portfolios based on the standards that were implemented. The underlined entries represent the cell into which the average of the panelists ratings would categorize that portfolio. For example, in Table 3 for the initial rating, Portfolio #5 was categorized as *Emerging* based on the implemented standards, but was given an average rating by the panelists that would have placed it in the *Progressing* category.

To calculate the means for each portfolio, portfolios classified as *Attempting* by panelists were assigned a score of 1, those classified as *Emerging* were assigned a score of 2, those classified as *Progressing* were assigned a score of 3, those classified as *Transitioning* were given a score of 4, and those classified as *Exceeding* were given a score of 5. Then all panelists' scores were averaged for each portfolio. To obtain the Category Averages (the data in the far right column in Tables 3-5), all of the panelists ratings for all portfolios in a performance level (based on the standards that were implemented) were averaged.

An examination of the data in Tables 3 through 5 indicates that, generally, the panelists did not classify the portfolios in the same manner as for the original standard setting. Additionally, the panelists tended to have varied ratings amongst themselves for many of the portfolios. For example, of the 42 portfolios evaluated, there was only one for which all of the panelists agreed on its classification.

Because the classification of individual portfolios may be viewed as a somewhat unreliable task, we aggregated the ratings of the panelists across the portfolios within each category as originally assigned. Examination of the column "Category Ave." in tables 3-5 indicates that the panelists, on average, assigned higher ratings to portfolios given higher classifications.

Although the average rating for all portfolios within a category increased as the performance level increased, panelists tended to disagree with the original ratings and with each other.

The lack of correspondence between the original classification and the panelists could result from one or more of several factors. The first possibility is that the original standards are in some sense "wrong" and the standards were simply set too high or too low. Another possibility is that during the standards validation meeting the intended rating task of the panelists was not done properly. This could have resulted from one of several possibilities including (but not limited to) lack of clarity of the task, the task being unrealistically complex and/or confusing, insufficient time for the task, and so on.

Because the panelists themselves tended to disagree with each other, even after discussion and exposure

to the original classification of the portfolios, it is less likely that the original standards are "wrong" (at least based on the evidence) than it is that the intended rating task was not done properly.

In support of this view, several panelists expressed confusion over the rating task and several spent a great deal of time trying to score the portfolios despite explicit directions not to do so. Additional evidence supporting the view that the panelists may have had a more difficult time than expected in rating the portfolios comes from the evaluation forms. In their evaluations panelists expressed a lack of clarity over several components of the validation meeting as well as, for Science, an insufficiency in the amount of time needed for the meeting. See Appendix B for a complete analysis of the evaluation forms.

Tasks Completed After the Standards Validation Meeting.

9. Analysis and Review of Panelists Ratings and Evaluations

Appendix B presents a summary of the evaluation forms completed by the panelists. These results provide information regarding anomalies in the procedures and validity of the process.

10. Prepare Report of the Meeting

The present report provides documentation of the procedures and results of the standards validation meeting in the establishment of performance standards for the MEA PAAP.

APPENDIX A

Powerpoint Presentation used at Opening Session

APPENDIX B

Evaluation Summary

What is your impression of the large group introductory session in the morning?

	N			Percent		
	Math	ELA	Science	Math	ELA	Science
Useful	8	8	4	61.5	57.1	33.3
Somewhat Useful	5	6	7	38.5	42.9	58.3
Not Useful	0	0	1	0.0	0.0	8.3

How clear were you with the performance level definitions?

	N			Percent		
	Math	ELA	Science	Math	ELA	Science
Very Clear	3	5	1	25.0	35.7	9.1
Clear	8	8	5	66.7	57.1	45.5
Somewhat Clear	1	1	4	8.3	7.1	36.4
Not at all Clear	0	0	1	0.0	0.0	9.1

How would you judge the length of this meeting?

	N			Percent		
	Math	ELA	Science	Math	ELA	Science
About right	11	12	2	91.7	85.7	16.7
Too little time	1	2	9	8.3	14.3	75.0
Too much time	0	0	1	0.0	0.0	8.3

What Factors influenced your ratings?

The performance level definitions

		N			Percent		
	Rating	Math	ELA	Science	Math	ELA	Science
Not at all Influential	1	0	0	0	0.0	0.0	0.0
	2	0	0	1	0.0	0.0	8.3
Moderately Influential	3	1	2	2	9.1	14.3	16.7
	4	3	3	3	27.3	21.4	25.0
Very Influential	5	7	9	6	63.6	64.3	50.0

The student work in the portfoilios

		N			Percent		
	Rating	Math	ELA	Science	Math	ELA	Science
Not at all Influential	1	0	0	0	0.0	0.0	0.0
	2	0	0	0	0.0	0.0	0.0
Moderately Influential	3	1	2	0	8.3	14.3	0.0
	4	2	4	5	16.7	28.6	41.7
Very Influential	5	9	8	7	75.0	57.1	58.3

Other panelists

		N			Percent		
	Rating	Math	ELA	Science	Math	ELA	Science
Not at all Influential	1	2	0	2	16.7	0.0	16.7
	2	3	7	3	25.0	50.0	25.0
Moderately Influential	3	6	6	6	50.0	42.9	50.0
	4	1	1	1	8.3	7.1	8.3
Very Influential	5	0	0	0	0.0	0.0	0.0

My experience in the field

		N			Percent		
	Rating	Math	ELA	Science	Math	ELA	Science
Not at all Influential	1	1	0	1	8.3	0.0	8.3
	2	1	0	0	8.3	0.0	0.0
Moderately Influential	3	1	5	6	8.3	35.7	50.0
	4	6	5	2	50.0	35.7	16.7
Very Influential	5	3	4	3	25.0	28.6	25.0

Other (Please specify)

		N			Percent		
	Rating	Math	ELA	Science	Math	ELA	Science
Not at all Influential	1	0	0	0	0.0	0.0	0.0
	2	1	0	0	16.7	0.0	0.0
Moderately Influential	3	1	1	1	16.7	20.0	50.0
	4	2	1	0	33.3	20.0	0.0
Very Influential	5	2	3	1	33.3	60.0	50.0

Do you believe the student portfolios were classified properly based on the initial standards?

	N			Percent		
	Math	ELA	Science	Math	ELA	Science
Definitely Yes	1	3	0	9.1	21.4	0.0
Probably Yes	10	6	4	90.9	42.9	40.0
Unsure	0	0	5	0.0	0.0	50.0
Probably No	0	5	1	0.0	35.7	10.0
Definitely No	0	0	0	0.0	0.0	0.0

Standard Setting for
Maine's Personalized Alternate Assessment Portfolio (PAAP)

In December, 2003 the first process related to standard setting for Maine's Personalized Alternate Assessment Portfolio (PAAP) was implemented. Because the PAAP Rubrics are so closely aligned with Maine's Learning Results Content Standards and each level is so explicitly described, psychometricians working with the state's assessment contractor indicated that the appropriate process was a validation of the standards outlined in the Rubrics.

As a result, a group of educators from around the state gathered in Augusta to participate in the validation process. The work was led by Measured Progress staff members. The day began with a general training session before people left to do their Content Area specific work which included a review of performance level definitions.

The work involved a systematic review of student work that Measured Progress staff had previously ordered by achievement level. When individual participants had completed the outlined process of review, there were panel discussions to reach consensus.

An analysis of the results indicated that the preponderance of ratings, using four levels of performance were adjacent rather than exact.

This validation process informed subsequent changes in the PAAP design and in the scoring process. Materials have been revised since to make the Rubrics more precise, and a common Task Bank of items has been developed for the use of students participating in state assessments through the PAAP.

Changes in other State Assessments led to the need for changes in the PAAP as we include students in grades 3, 5, 6, 7, and 10. These changes have now been made and will be implemented in 2006-2007. Following the scoring of PAAPs next spring, a second standard setting procedure involving personnel outside the Department of Education will be used to ensure that the PAAP is technically sound.

English Language Arts Content Area Panelists
2003 Validation Process

Chuck Anderson		Saco School District
Rich Leclair		Maine Youth Center
Louise St. Saviour	SAD 55	Hiram
Wayne Gammon	Union #44	Litchfield, Sabattus, Wales
Paige Coville	SAD 43	Rumford/Mexico
Laurie Lemieux		Monmouth School District
Terri Hicks		Windham School District
Leslie Abrams		Gorham School District
Gail Coughlan	SAD 11	Gardiner
Carla Baade Turner	SAD 6	Standish
Shari Casey	SAD 6	Standish
Barb Pineau	SAD 32	Ashland
Dawn Susee	SAD 33	Frenchville
Betsy Enright		Center for Community Inclusion University of Maine